

Exploiting High Dimensionality in Big Data

David Heckerman

Amazon

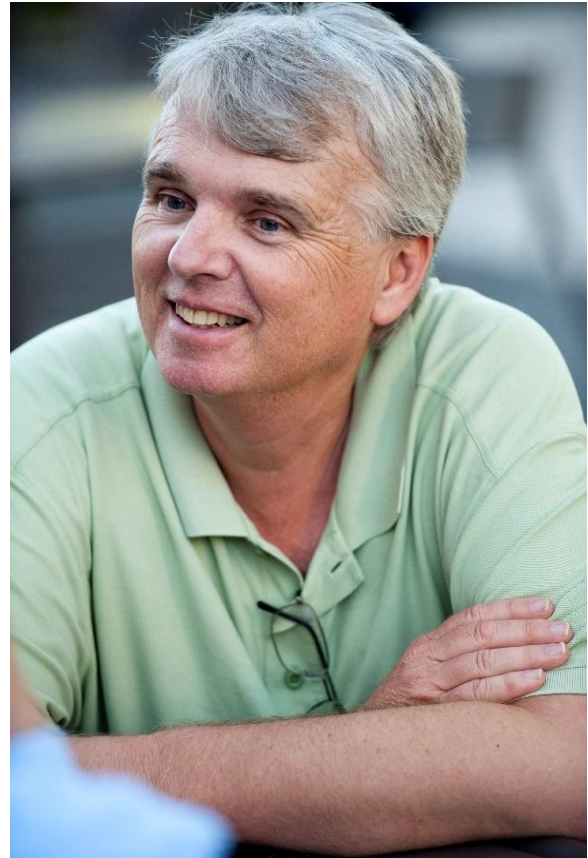
heckerma@amazon.com

ABSTRACT

There are two aspects of data that make them big: sample size and dimensionality. The advantages of large sample size have long been touted. In contrast, high dimensionality has typically been seen as an obstacle to successful analysis. In this talk, using the area of genomics as an example, I will illustrate some of the advantages of high dimensionality.

BIOGRAPHY

David Heckerman is a Distinguished Scientist and VP at Amazon. In his current scientific work, he is developing machine-learning approaches for biological and medical applications. In his early work, he demonstrated the importance of probability theory in Artificial Intelligence, and developed methods to learn graphical models from data including methods for causal discovery. While at Microsoft, he developed numerous applications including the junk-mail filters in Outlook, Exchange, and Hotmail, machine-learning tools in SQL Server and Commerce Server, handwriting recognition in the Tablet PC, text mining software in Sharepoint Portal Server, troubleshooters in Windows, and the Answer Wizard in Office. David received his Ph.D. (1990) and M.D. (1992) from Stanford University, and is an ACM, AAAI, and ACMI Fellow.



Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

KDD '19, August 4–8, 2019, Anchorage, AK, USA.

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6201-6/19/08.

DOI: <https://doi.org/10.1145/3292500.3340400>